社会关系网络中影响力用户发现方法研究*

何建民　殷　澍
(合肥工业大学管理学院　合肥 230009)
(过程优化与智能决策教育部重点实验室　合肥 230009)

摘要:【目的】为使结点投放的社会性网络广告绩效最大化，探索一种遴选网络影响力用户的新方法。【方法】从社会资本测量的视角，将社交网络上的用户间关系描述为社会关系网络基本图谱，据此建立用户影响力社会资本测度模型，通过对基本图谱下的用户随机游走行为发生可能性的概率计算，以识别和发现有影响力的用户。【结果】经识别的目标用户在其特定领域拥有相当的话语影响力，可对其他用户施加信息影响，经与"新浪微博"平台标识的影响力用户对比，验证该方法的有效合理性。【局限】未考虑社交网络用户的博文内容对其话语影响力的贡献。【结论】为企业投放社会性网络广告遴选有影响力用户，提供理论支持和实用方法。

关键词: 网络广告投放　社会关系图谱　用户社会资本测量　影响力用户发现模型
分类号: TP393　G35

1　引　言

在社交网络上，用户基于"利己"或"利他"的目的，以及社交和自我价值实现的需要，发布和传播特定话题的信息并与他人互动，以获得和增强其在网上的话语影响力[1]。拥有影响力的用户已成为企业投送社会化广告的重要营销资源。社会化网络广告与一般网络广告相比具有受众关注和互动影响的优势，绩效指标反馈直观。因此，发现和利用有影响力网络用户为企业推送广告以获得最大化的营销绩效便成为网络营销决策管理中的热点研究问题。

网络用户的话语影响力属于社会资本范畴[2]。Drushel[3] 认为意见领袖的社会资本就是话语影响力，它可促进对特定信息的传播，为企业投送网络广告。Lin[4]认为社会资本是内嵌于社会关系网络中的社会资源，行为人可获得和使用这些资源，包括权利、财富和声望等。由社会资本理论可知，社交网络用户之间彼此关注形成的社会关系是一种社会关系网络，用户可在该网络中发布和传播信息，以自己的信息选择方式和行为对他人施加信息的影响，从而获得和积累用户话语影响力[5]。网络用户的话语影响力社会资本资源由社会威望、专业背景和社交能力[6]等要素构成，体现为用户的个人属性和社会属性特征数据。

在现有研究用户影响力的文献和成果中，网上用户及其之间的信息关系被描述为不同类型的物理结构和关系网络，通过计算网上用户的特征属性数据，并经排序或综合计算，获得用户的影响力。代表性的成果主要有："用户博文内容–用户社交关系"二分图模型法[7-9]、用户自身影响力因素分析法[10-12]和用户间的多交互关系网络发现法[13-15]等。Zhu 等[7]、Ding 等[8]、Tang 等[9]利用"博文内容–用户"二分图关系，确定博文内容结点、用户结点及其子团之间的联接关系，发现网上用户博文传播的规律，并计算和发现影响力用户。Montangero 等[10]、Tarokh 等[11]、Liu 等[12]利用网络用户的个体行为特征或综合评价特征属性数据，计算用户的属性数据特征值，从中发现影响力用户；丁兆云等[13-14]将用户之间的信息"阅读"、"转发"、"回复"和"复制"行为关系描述为单关系和多关系的网络结

通讯作者：何建民，ORCID：0000-0002-9051-6929，E-mail：hejianmin@hfut.edu.cn。

*本文系国家自然科学基金重大项目"面向大数据的商务分析与计算方法以及支撑平台" (项目编号：71490725)和教育部人文社会科学研究规划基金项目"面向互联网大数据的舆论领袖话语影响力评估和治理方法研究" (项目编号：14YJA630015)的研究成果之一。

构，以 PageRank 算法计算用户随机游走下的行为发生概率，以找出影响力用户；齐超等[15]针对用户之间的信息"转发"、"评论"和"提及"行为关系，对用户的行为关系属性进行赋权，再通过叠加计算其行为特征数据，从中发现影响力用户。

综上，学者们采用多种理论方法以不同的视角，探索构成用户影响力的诸多属性及其特征，并将用户之间的信息关注关系映射为物理结构和关系网络，在计算用户的影响力时，主要是依据用户在网络结构中的位置和行为特征数据，将一种或多种用户个人属性或网络属性特征数据进行叠加和综合计算，以获得有影响力的用户。事实上，在社交网上，用户的影响力形成和增强与其社会关系演化过程相伴而生，用户影响力效用不仅与其自身的个人属性特征数据相关，还与其个人属性与社会属性特征数据共同作用因素相关。用户的影响力由其自身的个人属性和社会属性等多属性特征数据的共同作用而呈现影响力功效。本文以社会资本理论视角，将社交网上用户的关注行为关系描述为社会关系网络，将用户之间的信息"浏览"、"回复"和"分享"行为关系描述为单社会关系图谱，并确定用户的信息互动能力为用户影响力社会资本因素；在用户的单社会关系网络图谱下，将用户的个人属性与其信息交互行为属性特征数据共同作用关系，描述为多社会关系图谱，并确定用户的信息发布能力为又一用户影响力社会资本因素。设计用户社会关系网络图谱下的随机游走信息选择行为发生概率计算方法，并对用户的多关系属性下的特征数据采用乘积计算和结果综合排名的方法，获得社交网上的影响力用户，从而形成一套基于社会资本的社交网有影响力用户发现方法。

2　用户的社会关系网络图谱及其影响力发现模型

2.1　用户社会关系网络基本图谱

在社交网络中，假定 T 主题信息下的用户结点 i 发布或传播了一条有价值的信息，其他用户(如结点 j)会关注到结点 i 的信息，并以"浏览"、"回复"和"分享"行为与结点 i 进行互动，彼此形成关注关系(简称关系)，由此构成社会关系网络。事实上，用户结点 i 和结点 j 之间的关注关系有三种形式：

(1) 浏览。结点 i 发布的信息文本被结点 j"浏览"，结点 i 和结点 j 建立"浏览"关系；

(2) 回复。结点 j 使用提示标签"回复"了结点 i 所发出的信息文本，结点 i 和结点 j 建立"回复"关系；

(3) 分享。结点 j 转发了结点 i 所发出的信息文本，结点 i 和结点 j 建立"分享"关系。用户彼此关注而构建的社会关系网络基本图谱，如图 1 所示：

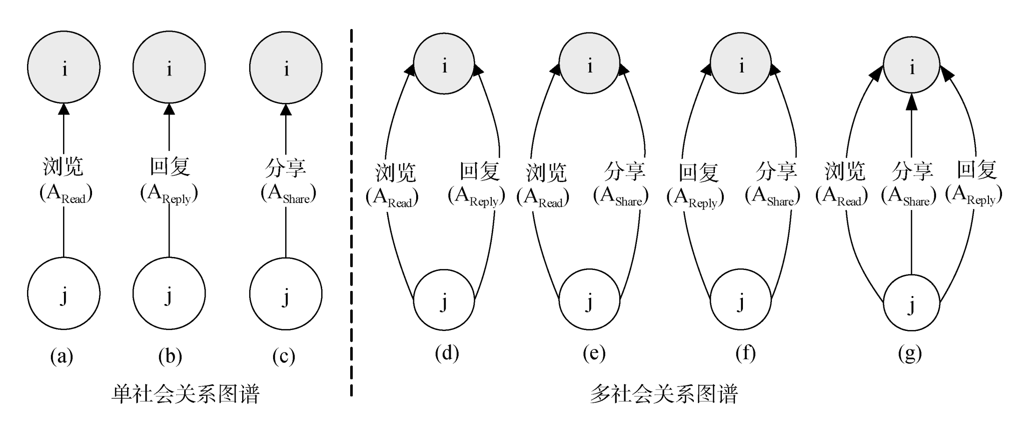


图 1　用户社会关系网络基本图谱

2.2　用户的社会关系状态转换图谱

在社交网络中，假定任意两个用户之间形成一种或多种社会关系，用户关注信息的行为结果一定是图 1 中的某一种状态形式。在图 1 中的(a)、(b)、(c)，为了描述用户结点 i 与结点 j 在既有的社会关系状态下向另一种新的关系状态转变的可能性，本文采用概率 α_{Read}、α_{Reply}、α_{Share} 分别表示结点 j 在任意两种社会关系状态之间进行行为变换的可能性。如用户结点 j"浏览"了结点 i 所发出的信息，可知结点 i 和结点 j 处于"浏览"状态，当结点 j 从"浏览"状态转变为"回复"状态时，结点 j 的社会关系状态发生改变的可能性用 α_{Reply} 表示，如图 2 中(a)。同理，由图 2 不同状态可知结点 j

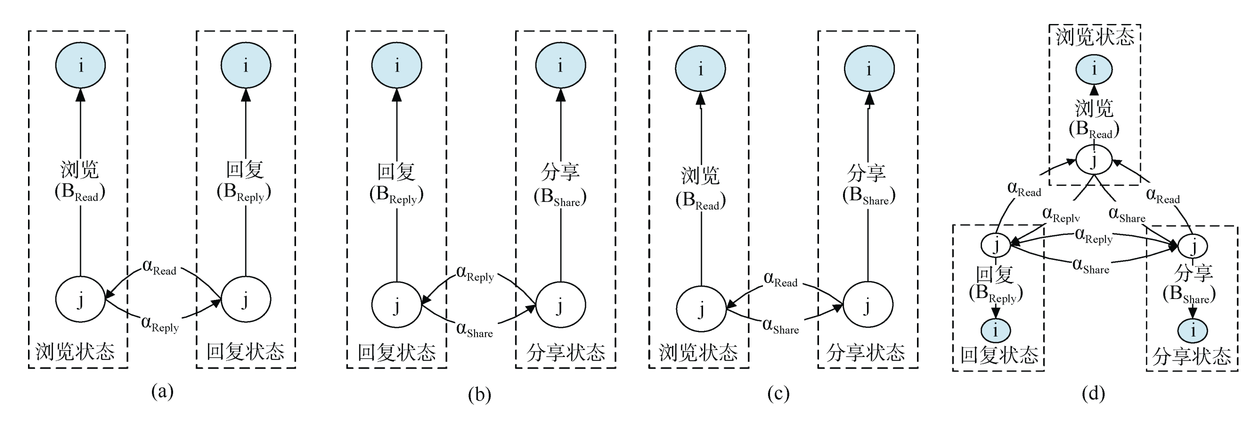


图 2　用户社会关系状态下的行为转换图谱

的关系转换概率α_{Read}，α_{Share}。

图 2 描述了用户结点在不同社会关系状态下的关系转换行为。可知，结点 j 对结点 i 提供的信息感兴趣，或认为该信息有价值才会发生社会关系转换。即结点 i 的信息对结点 j 具有利用价值，并对结点 j 施加了话语权影响，也即结点 i 对结点 j 具有话语影响力。当一个用户结点的话语影响力逐渐增大时，可认为该结点与 n 个其他用户结点形成了多种社会关系。左文明等[16]研究发现，用户的话语影响力与相关信息发布和社交关系能力成正相关性。因此，用户的影响力由其信息发布和社交关系能力确定，分别用信息发布能力和信息互动能力进行描述。其中，用户的信息发布能力是指用户运用自身的社会资源，发布带有观点倾向性的有价值信息，以对他人施加信息影响的能力，该能力可由用户所发信息的频度和被他人关注的强度计算获得；用户的信息互动能力是指用户结点与其他结点频繁地社交互动交流，以对他人施加信息影响的能力，该能力由用户之间的社交行为关系权重来计算。用户的话语影响力是其信息发布能力和信息互动能力共同作用的结果，用户的上述能力值相乘得到其话语影响力。用户的话语影响力程度越高，表明用户与他人搭建社会关系的可能性越大。如结点 j“浏览”了结点 i 的信息，其社会关系发生的可能性用“j 到 i 的社交行为发生概率”表示，可以通过计算结点 j 与其所有互动的结点的话语影响力值之和，再将该数值与结点 i 的话语影响力数值相比，得到结点 j 到结点 i 的社交行为发生概率。

图 1 为任意两个用户之间建立的社会关系网络图谱，据此可根据 PageRank 算法规则计算，即允许用户在信息选择行为中以随机游走方式选择任意一种行为的特性[17]，计算用户结点的社交行为发生融合概率，以获得结点在某一确定状态下的话语影响力。对于整个社交网来说，对每个用户结点的话语影响力复合值综合排名后，就能够得到拥有不同话语影响力的所有用户结点，从而可发现那些适合为企业投放网络广告的、营销目标与其话语影响力相配称的影响力用户。

2.3　用户社会关系网络的多关系图谱

在 T 主题信息下，假定结点用户的集合为 V，V 发布的文本集合为 C，V 中用户彼此存续一种或多种社会关系，则整个社交网中的结点用户形成一张庞大的社会关系网络图谱 G，G=(V, E)，其中 E 代表 G 中任意两个结点之间的关系集合。图 1 是任意两个用户结点之间存在关注关系的所有情形，由此可见图谱 G_f，$G_f=(V, E_{Read}\cup E_{Reply}\cup E_{Share})$。例如在图 1 的(a)中，任意结点之间存续“浏览”关系，用图谱 $G_{Read}=(V_{Read}, E_{Read}, W(E_{Read}))$ 表示，V_{Read} 是所有的存续“浏览”关系结点的集合，E_{Read} 是结点之间存续“浏览”关系的有向边集合，$W(E_{Read})$表示结点之间有“浏览”关系的权重，该权重体现为该结点对其他结点的信息互动能力。同理可知，任意两个结点处于“回复”关系下的图谱为 G_{Reply} 和处于“分享”关系下的图谱为 G_{Share}。且 $G\supseteq G_f\supseteq G_{Read}, G\supseteq G_f\supseteq G_{Reply}, G\supseteq G_f\supseteq G_{Share}$。

在社交网络中，由于结点用户与其他任意结点发生“浏览”关系是随机的，并没有标签标识，所以结点在该社会关系下的信息互动能力存在不确定性，导致无法计算结点在“浏览”关系下的权重；而用户结点与其他结点发生“回复”和“分享”关系时，其结果又是确

定的，可由信息标签确定其是“回复”还是“分享”关系。因此在计算“回复”和“分享”关系下的用户结点信息互动能力时，可根据结点的“回复”和“分享”次数，分别计算 $W(E_{Reply})$和 $W(E_{Share})$。

2.4　基于社会资本的用户影响力发现模型

基于社会资本的用户影响力发现模型如图 3 所示：

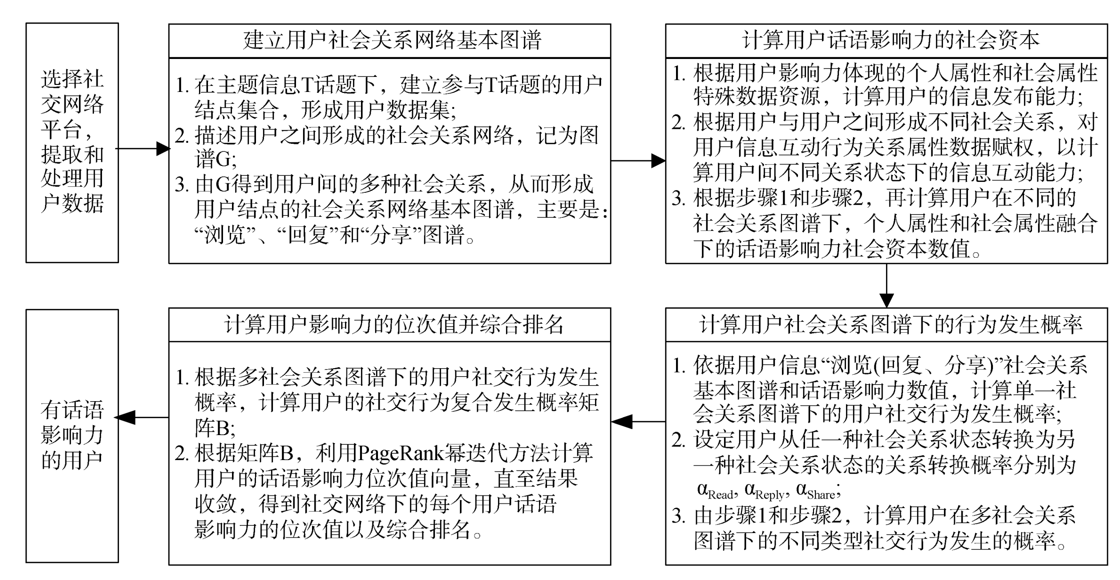


图 3　基于社会资本的用户影响力发现模型

该模型揭示了用户之间社会关系发生与转换变化的规律，由此可根据用户的社交行为方式随机游走，综合计算用户的社交行为发生概率及其关系转换概率，得到用户结点的话语影响力的位次数值和在整个社交网络中的综合排名，从而发现网上有影响力用户。

3　用户话语影响力的社会资本度量

3.1　用户的信息发布能力测度计算

在社会关系网络中，用户的话语影响力就是社会资本。用户影响力的获得和增强需要其持续发布和传播有价值的信息文本，以吸引他人的关注；同时，他还要与关注他信息的其他用户进行信息互动，由此才能对他人施加信息影响[18]。本文将用户结点所发信息文本被他人评论的数目和转发数相加再经平均，将其定义为用户结点的信息受关注程度；将用户结点在单位时间内所发出的信息文本数定义为结点的信息活跃程度，由此计算获得用户的信息发布能力。

(1) 用户的信息受关注程度测度计算

假设用户结点 i 在时间 t 内发出的信息文本数为 n_i，i 的 n_i 个信息文本被评论的总数为 c_i，被转发的总数为 r_i，则用户结点的信息受关注程度为 q_i，计算如公式(1)所示：

$$q_i = \frac{r_i + c_i}{n_i} \tag{1}$$

(2) 用户的信息活跃程度测度计算

假设用户结点 i 在时间 t 内发布的信息文本数为 n_i，用户结点的信息活跃程度为 a_i，则：

$$a_i = \frac{n_i}{t} \tag{2}$$

(3) 用户的信息发布能力测度计算

根据公式(1)和公式(2)计算可得用户结点的信息发布能力 d_i，如公式(3)所示：

$$d_i = q_i \cdot a_i \tag{3}$$

3.2　用户的信息互动能力测度计算

(1) 在用户“浏览”关系下

用户结点 i 参与了 T 主题话题信息的讨论，其信息一经发布即被 G 中用户结点所浏览，表明结点 i 所发布的信息能引起他人兴趣并有价值，因而结点 i 可能由此受到他人关注。他人在浏览结点 i 所发出的信息频度越高，表明结点 i 信息对他人施加的影响力越大。由 $G_{Read}=(V_{Read}, E_{read}, W(E_{Read}))$可知，用户结点 i 的信息互动能力为 w_{Read}。

用户结点 j“浏览”了结点 i 的信息，表明结点 j 体验了结点 i 所发信息的价值，并受到该信息的影响，在

过程中结点 j 用户有两类情形:

①结点 j 缺乏对 T 信息的了解和参与讨论必需的专业知识, 或结点 j 初入到社交网参与 T 话题讨论而受结点 i 信息的影响;

②结点 j 对 T 主题信息了解并有自己的见解, 愿意参与 T 主题信息的讨论, 同时也愿意发表独立见解。

在"浏览"社会关系基本图谱下, 区分上述两类用户的方法是: 先假定 G 下所有用户结点的信息发布能力阈值为 δ, 任意用户结点的信息发布能力为: $d_i,(i=1,\cdots,n)$, $\forall d_i \geqslant \delta$ 时, 可将结点 i 归属于第②类型的用户; $\forall d_i \leqslant \delta$ 时, 可将结点 i 归属于第①类型的用户。

针对于第①类型的用户 j, 若结点 j"浏览"了结点 i 的信息, 并将结点 i 加注为"好友"进行关注, 其决定因素源于结点 i 的"受欢迎程度"和结点 i 所"发布信息文本的数量"。假定用户结点 j 浏览了结点 i 所发出的信息, 可将结点 i 的信息互动能力定义为:

$$w_{Read}(v_j \to v_i) = \mu_i \times n_i \tag{4}$$

其中, μ_i 表示结点 i 的受欢迎程度, 可用结点 i 的"粉丝数/关注数"计算得到。

针对第②类型的用户 j, 结点 j 对好友结点 i 所发的与 T 主题相关的信息感兴趣。在公式(4)的基础上, 还要考虑用户结点 j"浏览"其好友结点 i 所发的信息与 T 主题信息内容的聚焦程度。用户结点 i 所发布的信息与 T 主题信息的聚焦程度越高, 表明其他用户如结点 j"浏览"结点 i 的信息文本可能性越大。假定用户结点 i 任意发布一则信息文本 $p_i(s)$, 则该信息文本与 T 主题信息的贴近程度为$\beta_i(s)$, 根据关键词 TF-IDF 检测技术和文本余弦相似度[19]可计算$\beta_i(s)$。假设结点 i 所发的信息文本与 T 主题信息的贴近程度阈值为θ, 当$\beta_i(s) \geqslant \theta$时, 则表明结点 i 所发信息文本 $p_i(s)$与 T 相关, 反之则无关。计算用户结点 i 在 t 时间内所发布的与 T 相关的文本总数, 将其与该时间段内所发文本总数相比, 其比值λ_i即为该用户结点 i 信息的聚焦程度。由此可得, 在"浏览"关系下的用户结点 i 的信息互动能力定义为:

$$w_{Read}(v_j \to v_i) = \mu_i \times n_i \times \lambda_i \tag{5}$$

(2) 在用户"回复"和"分享"关系下

用户结点之间存在"回复"和"分享"社会关系, 可通过标签检测予以确认。在"回复"关系图谱 $G_{Reply}=(V_{Reply},E_{Reply},W(E_{Reply}))$ 和"分享"关系图谱 $G_{Share}=(V_{Share},E_{Share},W(E_{Share}))$ 下, 用户结点 i 的信息互动能力 $w_{Reply}(v_j \to v_i)$ 和 $w_{Share}(v_j \to v_i)$, 由其所关注的其他用户结点"回复"或"分享"所发布的文本信息总数确定, 当用户结点 j 与其好友结点 i 有"回复"和"分享"关系时, 结点 i 的信息互动能力在"回复"下为: $w_{Reply}(v_j \to v_i)$, 在"分享"下为: $w_{Share}(v_j \to v_i)$。

3.3 用户话语影响力的社会资本计算

在图 2 中, 用户结点的话语影响力由其自身的信息发布能力 d 和信息互动能力 w 共同决定, 其中 w 是 w_{Read}、w_{Reply}、w_{Share} 的综合。结点 i 的话语影响力计算方法是对 d 和 w 两项数值的乘积。假定结点 i 和结点 j 是图 1 关系下的任意一种情形, 结点 i 的话语影响力 F, 可通过公式(6)–公式(8)计算得到用户话语影响力的社会资本测度:

$$F_i(Read) = w_{Read}(v_j \to v_i) \cdot d_i \tag{6}$$

$$F_i(Reply) = w_{Reply}(v_j \to v_i) \cdot d_i \tag{7}$$

$$F_i(Share) = w_{Share}(v_j \to v_i) \cdot d_i \tag{8}$$

4 计算用户的社交行为发生概率

4.1 单一社会关系

在图 1 的(a)、(b)、(c)中, 计算用户结点的话语影响力后, 要考察用户结点在图 1 的(d)、(e)、(f)、(g)下的结点社交行为发生的可能性, 主要是计算结点的社交行为发生概率。假定用户结点 j 在"浏览"了结点 i 的信息以后, 又转去"浏览"其他结点的信息, 则建立了一种新的多边关系, 即"浏览–浏览"关系。令该结点 j"浏览"他人信息行为发生的概率矩阵为 A_{read}, 则结点用户的"浏览"、"回复"、"分享"行为发生概率为:

(1) 任意用户结点 v_j 随机"浏览"结点 v_i 信息文本的社交行为发生概率为:

$$A_{read}(v_j \to v_i) = \frac{F_i(Read)}{\sum_{v_{j\varepsilon v}} F_v(Read)} \tag{9}$$

(2) 任意用户结点 v_j 随机"回复"结点 v_i 信息文本的社交行为发生概率为:

$$A_{Reply}(v_j \to v_i) = \frac{F_i(Reply)}{\sum_{v_{j\varepsilon v}} F_v(Reply)} \tag{10}$$

(3) 任意用户结点 v_i 随机"分享"结点 v_j 信息文本的社交行为发生概率为:

$$A_{Share}(v_j \to v_i) = \frac{F_i(Share)}{\sum_{v_{j\notin v}} F_v(Share)} \quad (11)$$

其中，$v_j \notin v$，v 代表被 v_j "浏览"，或"回复"，或"分享"的所有结点。

4.2　多社会关系

在图1关系下，除去结点以(a)、(b)、(c)形式与其他结点进行社交互动外，还能以(d)、(e)、(f)、(g)形式与其他结点进行社交互动。因此，需要建立一种新型社会关系，即为多社会关系图谱。

假定结点用户的社会关系状态属于图1中(a)、(b)、(c)中的任意一种，则设结点从任意一种关系状态转换为"浏览(回复、分享)"关系状态的社会关系转换概率分别是：α_{Read}、α_{Reply}、α_{Share}，且满足$\alpha_{Read}+\alpha_{Reply}+\alpha_{Share}=1$。结点发布信息除在图1中(a)、(b)、(c)下，以随机游走方式对他人施加话语影响力外，还能以此为起点与其他任意结点构建新的社会关系，如图1中(d)、(e)、(f)、(g)。由此可计算，在既定的社会关系图谱下结点随机游走下步行为变化的可能性，再与结点发生社会关系状态转换的可能性进行综合，便得到该结点对其他结点施加话语影响力的所有可能性，即计算多社会关系下结点社交行为发生概率。

在T主题话题下，假定γ是结点在图1中(a)、(b)、(c)中与其他任意结点建立社会关系的可能性。由图2所示的多关系图谱可知，结点间的社交行为发生概率是：B_{Read}、B_{Reply}、B_{Share}，这时考虑结点在图1中的社交行为发生可能性A_{Read}、A_{Reply}、A_{Share}，同时需要考虑结点在图2的多社会关系状态转换发生的可能性α_{Read}、α_{Reply}、α_{Share}的概率计算方法。

(1) 用户结点j在"浏览" i后的社交行为发生概率$B_{Read}(v_j \to v_i)$，得到：

$$B_{Read}(v_j \to v_i) = \alpha_{Read} \times \left[(1-\gamma)\times A_{Read}(v_j \to v_i) + \frac{\gamma}{n}\right] \quad (12)$$

(2) 用户结点j在"回复" i后的社交行为发生概率$B_{Reply}(v_j \to v_i)$，得到：

$$B_{Reply}(v_j \to v_i) = \alpha_{Reply} \times \left[(1-\gamma)\times A_{Reply}(v_j \to v_i) + \frac{\gamma}{n}\right] \quad (13)$$

(3) 用户结点j在"分享" i后的社交行为发生概率$B_{Share}(v_j \to v_i)$，得到：

$$B_{Share}(v_j \to v_i) = \alpha_{Share} \times \left[(1-\gamma)\times A_{Share}(v_j \to v_i) + \frac{\gamma}{n}\right] \quad (14)$$

其中，参与T主题话题互动的所有的用户结点总数为n。

5　用户的话语影响力复合计算和综合排名

在T主题话题信息下，假定参与T的用户结点数为n，这些用户的话语影响力位次值所组成的向量为P。P中任意一个用户结点i的话语影响力位次值为$p(v_i)$。由PageRank算法[17]可知，计算$p(v_i)$值即在图2中计算结点i的追随者的排名位次值，再将所有与结点i有社会关系的追随者的排名位次值与其社交行为发生概率相乘，再逐个累加得到结点i的话语影响力总位次值，$p(v_i)$计算公式如下：

$$p(v_i) = \left[\sum_{v_{i,j}\in E_{read}} B_{Read}(v_j \to v_i)\,p(v_j) + \sum_{v_{i,j}\in E_{reply}} B_{Reply}(v_j \to v_i)p(v_j) + \sum_{v_{i,j}\in E_{share}} B_{Share}(v_j \to v_i)p(v_j)\right] \quad (15)$$

其中，将$p(v_j)$从括号中移出，括号内的公式恰好是B_{Read}、B_{Reply}、B_{Share}行为发生概率矩阵转置相加，再与P向量矩阵相乘所得的第i行的$p(v_i)$数值。

假设在图2中，用户结点i行为发生概率复合矩阵为B，$B=B_{Read}+B_{Reply}+B_{Share}=1$。根据PageRank幂迭代计算方法，可计算n个结点的话语影响力的排名位次值，得到其排名位次值向量矩阵P；再根据马尔科夫链及其有限性定理[20]，可知该向量矩阵P的结果收敛且唯一。

$$P_k = B^T P_{k-1} \quad (16)$$

其中，k的初始值为0，初始位次值向量令为P_0，$P_0 = (1,1,\cdots,1,1,)^T$。$P_k, P_{k-1}$是经过第k次和第k–1次迭代计算以后所获得的n个结点话语影响力位次值的向量。

6　实验及结果讨论

6.1　实验平台和数据采集

实验平台为"新浪微博"社交网络平台，其数据采集采用新浪微博提供的 API 编程接口通过程序自动获得，数据采集的时间为 2014 年 10 月 1 日–2014 年 10 月 31 日。在获取的用户数据包中，挑选用户讨论话题"#爸爸去哪儿了#"作为 T 主题话题信息，采集的微博用户结点数据，主要包括结点自身特征属性数据、互

动行为关系属性数据，以及结点发布和传播的博文及其互动、转发的日常记录数据。在剔除无效的 ID 用户结点号后，获得参与 T 主题话题信息发布、浏览、回复和转发的用户总数为 106，博文信息记录总数为 1 301。用户数据记录属性项包括：用户昵称(A0)、是否加 V(A1)、关注数(A2)、粉丝数(A3)、总发微博数(A4)，以及在数据采集周期内的微博数(A5)和微博被评论和被转发数之和(A6)。将上述微博用户数据集合导入基本信息数据库，部分数据属性项列表如表 1 所示。同时，根据这些用户彼此之间的关注关系，以“0-1”矩阵形式建立结点用户关注关系数据库，如表 2 所示。

表 1　微博用户基本信息数据库(部分)

WEIHAO_ID	A0	A1	A2	A3	A4	A5	A6
2580538964	有点任性的射手	0	185	129	366	10	1
1795128540	杨仔 Melody	0	285	531	1 511	14	64
1900441543	丽–歌	0	291	1 378	5 994	100	535
2127830227	混在阳光下丶T	0	794	229	1 116	450	3
3185540783	萍萍萍萍平	0	27	8	13	1	1
…	…	…	…	…	…	…	…

表 2　微博结点用户关注关系数据库(部分)

ID	2308923167	1936123170	1367732612	2020410163	3816239594	…
2308923167	0	0	0	0	1	…
1936123170	1	0	1	0	0	…
1367732612	0	1	0	1	1	…
2020410163	1	0	0	0	1	…
3816239594	1	1	1	1	0	…
…	…	…	…	…	…	…

6.2　实验步骤和结果讨论

(1) 微博用户结点自身的信息发布能力计算

在微博用户基本信息数据库中，由公式(1)–公式(3)可计算微博用户的信息发布能力 d。如昵称为“菜菜_家有小小菜”的用户，他的信息发布能力值为 2.6777，经计算可得所有的微博用户信息发布能力。假定信息发布能力阈值为δ=1，由微博结点的信息发布能力是大于或小于δ判定条件可知，该微博结点属于 3.2 节所定义的第①或第②类用户类型，由此可计算微博用户结点信息发布能力。计算结果如表 3 所示：

表 3　微博用户信息发布能力计算结果和用户类型(部分)

WEIHAO_ID	A0	A1	A2	A3	A4	A5	A6	d	type
2308923167	UNES 优理氏	0	1 670	22 296	4 084	150	40	1.2903	②
1936123170	妆–只是一种态度	0	555	285	9 871	500	50	1.6129	②
1367732612	油切麦兜	0	222	654	29	3	1	0.0323	①
2020410163	洛丽塔奇奇大王 T	0	476	655	167	6	1	0.0333	①
3816239594	发乱女	0	283	104	306	10	9	0.2903	①
1817474694	菜菜_家有小小菜	0	256	231	1083	45	83	2.6777	②
…	…	…	…	…	…	…	…	…	…

(2) 图1微博用户社交行为发生概率矩阵计算

根据微博结点用户关注关系数据库，由其用户类型判定条件和计算公式(4)–公式(6)和公式(9)，可计算微博用户在“浏览”关系下的社交行为发生概率。计算结果如表 4 所示：

表 4　微博用户在“浏览”关系下的社交行为发生概率(部分)

ID	2308923167	1936123170	1367732612	2020410163	3816239594	…
2308923167	0	0	0	0	0.0015	…
1936123170	0.9973	0	0.0004	0	0	…
1367732612	0	0.0003	0	0.073	0.012	…
2020410163	0.0573	0	0	0	0.002	…
3816239594	0.326	0.0128	0.0154	0.0171	0	…
…	…	…	…	…	…	…

同理，可获得微博用户在“回复”和“分享”关系下的社交行为发生概率。

(3) 图 2 中的微博用户社交行为发生概率及其矩阵计算

由步骤(2)计算可得图 1 中的用户社交行为发生概率，再利用公式(12)–公式(14)计算微博用户在图 2 下的社交行为发生概率。先设定结点的社会关系转换概率为α，微博用户“浏览”了好友用户博文，即认为该微博用户受到好友的信息影响，但影响不及微博用户回复或转发其好友博文而受到其影响作用大[13]，可得，$\alpha_{Read}<\alpha_{Reply}$，$\alpha_{Read}<\alpha_{Share}$。设$\alpha_{Read}$、$\alpha_{Reply}$、$\alpha_{Share}$分别为 0.2、0.4、0.4，根据经验值将γ设为 0.15。由此计算结点在图 2 中的社交行为发生概率 B_{Read}、B_{Reply}、B_{Share}，以计算 B_{Read} 为例，如表 5 所示：

表 5　B_{Read} 下的结点间行为发生概率(部分)

ID	2308923167	1936123170	1367732612	2020410163	3816239594	…
2308923167	0.00028	0.00028	0.00028	0.00028	0.00051	…
1936123170	0.14988	0.00028	0.00034	0.00028	0.00028	…
1367732612	0.00028	0.00033	0.00028	0.00138	0.00208	…
2020410163	0.00888	0.00028	0.00028	0.00028	0.00058	…
3816239594	0.04918	0.0022	0.00259	0.00285	0.00028	…
…	…	…	…	…	…	…

同理，得到 B_{Reply}、B_{Share} 的结点间社交行为发生概率。

(4) 微博用户的话语影响力位次值与综合排名

由步骤(3)和公式(15)、公式(16)计算微博用户的话语影响力位次值与综合排名，如表 6 所示：

表 6　微博用户的话语影响力位次值和综合排名(部分)

综合排名	ID	A0	A1	A3	位次值
1	3845175940	湖南卫视爸爸去哪儿	0	799	0.561119259
2	3817825983	小小志 kimi 粉丝童乐会	0	258 746	0.462998398
3	1843633441	搜狐视频	1	2382 073	0.351090861
4	1864391637	乐视综艺	1	40 473	0.269427536
5	2120293981	Holly 大娟	0	819	0.264906412
6	1843633441	搜狐娱乐	1	960 717	0.227777091
7	1068803873	自由主妇	1	338 292	0.195439281
8	2020410163	洛丽塔奇奇大王 T	0	655	0.164923929
9	1779850265	爱奇艺娱乐	1	2 034 084	0.156936338
10	1565752662	赫呵	0	4 844	0.103160867
…	…	…	…	…	…

根据步骤(2)–步骤(4)可计算获得微博用户的话语影响力的综合排名，由此可发现在 T 主题话题下有话语影响力的微博用户。假定某企业的营销主题内容与该 T 主题话题所讨论的内容相一致，据此可在表 6 中，帮助企业推荐排名靠前的用户，作为企业投送 T 主题广告的备选影响力用户，从而为企业的社会化营销广告投放提供理论支持和实用方法。

由表 6 所得的微博用户话语影响力结果可知，本文提供的微博用户话语影响力发现方法不同于新浪微博官方提供的微博用户影响力推荐方法。“新浪微博”平台是以微博的粉丝数或用户是否‘加 V’评判用户影响力。事实上，微博用户归属于哪个社交圈是由其深谙的领域知识、兴趣和能力因素决定，判定用户的影响力效用需要综合考虑他在社交圈中的网络威望、专业背景和社交能力等要素。表 6 给出用户的话语影响力排名彰显其在上述三个方面的擅长和能力，可为企业选择用户的擅长领域和能力提供决策支持。由此可帮助企业广告传播，使其广告便于目标受众关注、产生兴趣、搜索和行动，促使其购买和分享价值形成口碑。探索利用用户话语影响力因素与广告绩效因素的关系选择合适的广告投放用户方法，使企业广告绩效最大化是一种新的尝试。

7 结 语

社会性网络及其用户的话语影响力是企业网络营销重要的战略资源，借助用户的话语影响力可为企业投放社会性网络广告，以获取比一般网络广告更优的推广效率和效益。如何在社交网上发现有话语影响力的用户是社会化营销绩效最大化领域中的热点研究问题。

本文以社会资本理论视角，研究社交网络上的用户影响力及其影响力效用的形成与增强过程，探索用户之间的信息选择行为和关系、用户社会关系网络和用户的社会资本形成规律，旨在探明社交网上用户话语影响力的社会资本要素构成，发现用户的信息发布能力和信息互动能力是其社会资本的主要内容，由此定义了用户话语影响力社会资本的个人属性、社会属性及其行为关系属性的特征数据，并采用概率论方法计算其影响力属性特征的相关数据，以确定社交网上的影响力用户，从而形成一套基于社会资本的社交网络影响力用户发现的方法。

研究表明，社交网络上用户间的社交关系其实是一种社会关系网络。用户的影响力由其信息发布能力和信息互动能力构成，它是用户社会资本资源的主要体现，由用户的个人属性和社会属性特征数据表征。建立了用户单关系和多关系的社会关系图谱，由此定义了用户的社交关系行为发生概率和随机游走行为与关系转换概率，以计算用户对其他用户施加信息影响行为发生的可能性(即影响力效用)，由此构建基于社会资本的影响力用户发现模型，描述不同社会关系下的用户社交行为发生概率和关系转换概率，并采用 PageRank 随机游走思想，计算用户话语影响力的社会资本位次值和综合排名，以发现有影响力的用户，最终形成一套基于社会资本的社交网影响力用户发现的方法，通过对“新浪微博”用户数据的进行实验，验证其方法的合理有效性。该方法可作为企业挑选有影响力的用户以投放社会化广告的方法，对帮助企业谋求最大化的社会化营销绩效具有重要的价值。下一步的研究工作将针对社交网上用户发布的信息内容，研究用户传播的信息与其影响力增长的贡献关系。

参考文献：

[1] 何建民，郑哲. 社交网络消费者抱怨“动机–行为”模型研究[J]. 合肥工业大学学报：社会科学版, 2015, 29(1): 1-9. (He Jianmin, Zheng Zhe. Study of “Motivation-Behavior” Model of Consumers’ Complaints on Social Network [J]. Journal of Hefei University of Technology: Social Science, 2015, 29(1): 1-9.)

[2] O’Connor L G, Dillingham L L. Personal Experience as Social Capital in Online Investor Forums [J]. Library & Information Science Research, 2014, 36(1): 27-35.

[3] Drushel B E. HIV/AIDS, Social Capital, and Online Social Networks [J]. Journal of Homosexuality, 2013, 60(8): 1230-1249.

[4] Lin N. Social Capital: A Theory of Social Structure and Action [M]. Cambridge University Press, 2002.

[5] 唐晓波，房小可. 融入社会关系的微博排名策略研究[J]. 现代图书情报技术, 2013(9): 74-81. (Tang Xiaobo, Fang Xiaoke. Research on Microblog Ranking Strategy with the Social Relations [J]. New Technology of Library and Information Service, 2013(9): 74-81.)

[6] Subbian K, Sharma D, Wen Z, et al. Finding Influencers in Networks Using Social Capital [C]. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in

Social Networks Analysis and Mining.2013: 25-28.
[7] Zhu Z, Su J, Kong L. Measuring Influence in Online Social Network Based on the User-Content Bipartite Graph [J]. Computers in Human Behavior, 2015, 52: 184-189.
[8] Ding Z, Wang H, Guo L, et al. Finding Influential Users and Popular Contents on Twitter [C]. In: Proceedings of the 16th International Conference on Web Information System Engineering, Miami, FL, USA. Springer International Publishing, 2015: 267-275.
[9] Tang B, Lu T, Gu H, et al. Measuring Domain-Specific User Influence in Microblogs: An Actor-Network Theory Based Approach [C]. In: Proceedings of the 19th International Conference on Computer Supported Cooperative Work in Design. IEEE, 2015: 314-319.
[10] Montangero M, Furini M. TRank: Ranking Twitter Users According to Specific Topics [C]. In: Proceedings of the 12th Conference on Consumer Communications and Networking, Las Vegas, NV, USA. IEEE, 2015: 767-772.
[11] Tarokh M J, Arian H S, Speily O R B. Discovering Influential Users in Social Media to Enhance Effective Advertisement [J]. Advances in Computer Science: An International Journal, 2015, 4(5): 23-28.
[12] Liu N, Li L, Xu G, et al. Identifying Domain-Dependent Influential Microblog Users: A Post-Feature Based Approach [C]. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence. AAAI, 2014: 3122-3123.
[13] Ding Z, Jia Y, Zhou B, et al. Mining Topical Influencers Based on the Multi-relational Network in Micro-blogging Sites [J]. Communications, 2013, 10(1): 93-104.
[14] 丁兆云，周斌，贾焰，等. 微博中基于多关系网络的话题层次影响力分析[J]. 计算机研究与发展，2013，50(10): 2155-2175. (Ding Zhaoyun, Zhou Bin, Jia Yan, et al. Topical Influence Analysis Based on the Multi-Relational Network in Microblogs [J]. Journal of Computer Research and Development, 2013, 50(10): 2155- 2175.)
[15] 齐超，陈鸿昶，于洪涛. 基于用户行为综合分析的微博用户影响力评价方法[J]. 计算机应用研究，2014，31(7): 2004-2007. (Qi Chao, Chen Hongchang, Yu Hongtao. Method of Evaluating Micro-blog Users' Influence Based on Comprehensive Analysis of User Behavior [J]. Application Research of Computers, 2014, 31(7): 2004-2007.)
[16] 左文明，王旭，樊偿. 社会化电子商务环境下基于社会资本的网络口碑与购买意愿关系[J]. 南开管理评论，2014 (4): 140-150. (Zuo Wenming, Wang Xu, Fan Chang. Relationship Between Electronic Word of Mouth and Purchase Intention in Social Commerce Environment: A Social Capital Perspective [J]. Nankai Business Review, 2014(4): 140-150.)
[17] Wu X, Kumar V. The Top Ten Algorithms in Data Mining [M]. Chapman and Hall/CRC, 2009.
[18] 毛佳昕，刘奕群，张敏，等. 基于用户行为的微博用户社会影响力分析[J]. 计算机学报，2014, 37(4): 791-798. (Mao Jiaxin, Liu Yiqun, Zhang Min, et al. Social Influence Analysis for Micro-blog User Based on User Behavior [J]. Chinese Journal of Computers, 2014, 37(4): 791-798.)
[19] Russell M A. Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More [M]. O'Reilly Media, 2013.
[20] Li L, Sun L, Ning G. Deterioration Prediction of Urban Bridges on Network Level Using Markov-Chain Model [J]. Mathematical Problems in Engineering, 2014, 9(1): 129-145.

作者贡献声明:

何建民：提出研究思路，设计研究方案，论文最终版本修订；
殷澍：采集、清洗和分析数据，进行实验；
何建民，殷澍：论文起草。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据见期刊网络版 http://www.infotech.ac.cn。
[1] 殷澍. #爸爸去哪儿#话题下微博用户发布、转发、回复博文情况基本数据.xlsx. #爸爸去哪儿#话题下微博用户发布、转发、回复博文情况基本数据.
[2] 殷澍. 微博用户基本信息表.xlsx. 微博用户基本信息表.
[3] 殷澍. 微博用户信息发布能力计算结果和用户类型数据.xlsx. 微博用户信息发布能力计算结果和用户类型数据.
[4] 殷澍. 微博结点用户关注关系邻接矩阵数据.xlsx. 微博结点用户关注关系邻接矩阵数据.
[5] 殷澍. 微博用户"浏览"关系下的社交行为发生概率矩阵数据.xlsx. 微博用户"浏览"关系下的社交行为发生概率矩阵数据.
[6] 殷澍. 微博用户"回复"关系下的社交行为发生概率矩阵数据.xlsx. 微博用户"回复"关系下的社交行为发生概率矩阵数据.
[7] 殷澍. 微博用户"分享"关系下的社交行为发生概率矩阵数据.xlsx. 微博用户"分享"关系下的社交行为发生概率矩阵数据.
[8] 殷澍. Bread 下的微博用户结点行为发生概率矩阵数据.xlsx. Bread 下的微博用户结点行为发生概率矩阵数据.
[9] 殷澍. Breply 下的微博用户结点行为发生概率矩阵数据.xlsx. Breply 下的微博用户结点行为发生概率矩阵数据.
[10] 殷澍. Bshare 下的微博用户结点行为发生概率矩阵数

据.xlsx. Bshare 下的微博用户结点行为发生概率矩阵数据.
[11] 殷澍. 矩阵 B.xlsx. 矩阵 B.
[12] 殷澍. 矩阵 B 转置.xlsx. 矩阵 B 转置.
[13] 殷澍. 幂迭代法计算微博用户位次值的迭代步骤和数值.xlsx. 幂迭代法计算微博用户位次值的迭代步骤和数值.
[14] 殷澍. 幂迭代主特征值法计算得到的微博用户排名及其位次值和数值.xlsx. 幂迭代主特征值法计算得到的微博用户排名及其位次值和数值.
[15] 殷澍. 数据逻辑关系注释.txt. 数据逻辑关系注释.

收稿日期: 2015-10-26
收修改稿日期: 2016-01-15

Identifying Influential Users in Social Networks

He Jianmin Yin Shu
(School of Management, Hefei University of Technology, Hefei 230009, China)
(Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei University of Technology, Hefei 230009, China)

Abstract: [**Objective**] This paper aims to identify the influential users in social network systems, which could help us maximize the online advertising effects. [**Methods**] First, we constructed the basic graphs to describe relationship among the social network system users from the perspective of social capital measurement. Second, we built the influence measurement model based on the newly constructed graphs. Finally, we identified the influential users by calculating the probabilities of users' random browsing behaviors. [**Results**] The proposed method could identify users with big online influence. They were more capable of affecting others in related fields than the influential users listed by the Sina Weibo. [**Limitations**] The proposed method did not evaluate the impacts of user-generated contents in social network systems while measuring the users' influence. [**Conclusions**] The proposed method could help business owners identify influential users in the social network system to improve the effectiveness of online advertisements.

Keywords: Online advertisement placement Social relationship graphs Social capital measurements of users Identification model of influential users